

User Behavior in Tagging in the OPAC: The Example of the Faculty of Humanities and Social Sciences Library in Zagreb

Aleksandra Pikić

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
apikic@ffzg.hr

Dorja Mučnjak

Faculty of Humanities and Social Sciences Library
Ivana Lučića 3, HR-10000 Zagreb, Croatia
dmucnjak@ffzg.hr

Summary

Paradigmatic change has occurred in understanding the relationship of Internet users towards the content available on the Web. Unlike the earlier role of users as passive consumers, the new network, Web 2.0 is based on the user-friendly technologies which are focused on collaborative and interactive information services, social bookmarking services, where users can categorize and store their own web links, images, bibliographic records or PDF files. These technologies allow users to independently organize information in the ways which suit them best. Unlike the traditional organizing, where information specialists or authors describe, organize and classify contents, the users of social bookmarking services create metadata themselves using uncontrolled keywords - tags.

Library of the Faculty of Humanities and Social Sciences in Zagreb also allows its users to tag bibliographic records of library materials. In June 2010 the option of adding and browsing tags in the Koha catalogue was plugged in, and the users were informed about this new library service.

The aim of this study is to explore how the Library users tag bibliographic records, i.e. what keywords are used to organize the information in the catalogue. The research problems were the following: which users are the most common taggers, which type of metadata is marked by tags and how do theme tags correspond to the key words in the catalogue?

Key words: social tagging, subject indexing, OPAC, Faculty of Humanities and Social Sciences Library in Zagreb, folksonomy

Introduction

By the mid 2000s a paradigmatic change has occurred in the understanding of the relationship of Internet users towards the content available on the Web. Unlike the earlier role of users as passive consumers, the new network, Web 2.0 is based on the user-friendly technologies which are focused on collaborative and interactive information services¹. Such technologies are social bookmarking services like Delicious², Flickr³ and CiteULike⁴, where users can categorize and store their own web links, images or PDF files. These technologies allow users to independently organize information important to them in the ways which suit them best. Unlike the traditional knowledge organization, where information specialists or authors describe, organize and classify contents, the users of social bookmarking services create metadata themselves using uncontrolled keywords - tags.

Tag can be defined as "any word that defines a relationship between the online resource and the concept in the user's mind"⁵. The process of adding tags is usually called collaborative or social tagging, and the result of this process is a folksonomy. Folksonomies have evolved from the above-mentioned social bookmarking systems. Munk and Moerk define folksonomies as taxonomies created by users who freely create descriptive metadata by tagging documents⁶. The English Wikipedia defines folksonomy as a classification system that has resulted from the collaborative creation and management of tags in order to capture and categorize content. The term was coined by the information architect Thomas Vander Wal in 2003. This neologism is composed of two terms: the term folk (people) and taxonomy, so the literal translation of the concept would be the taxonomy of the people. Folksonomy is not a classification but a flexible horizontal categorization, which consists of associative, but unrelated concepts that users can add and browse without professional supervision. It represents a departure from the traditional classification; the expert categorization of information through the use of a controlled vocabulary. Everybody can tag at any moment on the Internet regardless of language, time or place where they are. Unlike the subject indexing done by the experts, this tagging is more up to

¹ Macan, Bojan. Tehnologije Web 2.0 i njihova primjena u knjižnicama – iskustva Knjižnica Instituta Ruđer Bošković s posebnim osvrtom na njezin blog. // *Kemija u industriji*. 58, 5(2009); 226.

² Delicious. [quoted: 2010-09-15] URL: <http://www.delicious.com>

³ Flickr. [quoted: 2010-09-15] URL: <http://www.flickr.com>

⁴ CiteULike. [quoted: 2010-09-15] URL: <http://www.citeulike.org>

⁵ Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // *Library and information science research*, 32(2010); 192.

⁶ Munk, T.B. ; Moerk, K. Folksonomies, tagging communities and tagging strategies: an empirical study. // *Knowledge organization*, 34, 3(2007); 116.

date and captures changes when they happen⁷. In addition to these advantages, the tagging system also has serious flaws. Unlike the traditional classification systems and thesauri, there is no expert control over folksonomies, nor any selection criteria or instructions for tagging. All of this generates a large number of homonyms, terms in the singular as well as plural, “sloppy” tags such as misspellings⁸ and the danger of different users attributing different meanings to the same tags⁹.

Spiteri (according to Lu, Park and Hu) claims that a folksonomy can create added value to public library catalogues since users can organize their personal information space, create additions to the existing controlled vocabularies and create online communities of interests¹⁰.

Kakali and Papatheodorou report that, lately, museums, archives and libraries have also allowed their users to tag documents in their catalogues by themselves¹¹. Based on added tags in catalogues, the museums The Steve Collaboration and the Penn Museum have been investigating which items in their work of art collections their users consider important and significant. Furthermore, many libraries have upgraded their public web catalogues with the social tagging systems such as LibraryThing for Libraries¹², VuFind¹³, Scriblio¹⁴ etc., and in that manner they are enhancing their catalogues. The authors think that user tags can assist librarians in modernizing the vocabularies of their classification systems and that they can reduce the percentage of unanswered user queries.

The Faculty of Humanities and Social Sciences Library in Zagreb also enables its users to tag bibliographic records of the library collection. In June 2010, the option of adding and reviewing tags in the catalogue Koha was included, and the users were informed about the new service offered by the library (Image 1). In order for users to be able to add tags, they must register with their

⁷ Steele, Tom. The new cooperative cataloging. // *Library Hi Tech*, 27, 1(2009); 70.

⁸ Thomas, Marliese; Caudle, Dana M.; Schmitz, Cecilia. Trashy tags : problematic tags in LibraryThing. // *New Library World*, 111, 5/6(2010); 225.

⁹ Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // *Library and information science research*, 32(2010); 192.

¹⁰ Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congress Subject Headings. // *Journal of information science*, 36, 6(2010); 767.

¹¹ Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // *Library and information science research*, 32(2010); 192.

¹² LibraryThing for Libraries. [quoted: 2011-05-15] URL: <http://www.librarything.com/forlibraries>

¹³ VuFind. [quoted: 2011-05-15] URL: <http://vufind.org/>

¹⁴ Scriblio. [quoted: 2011-05-15] URL: <http://about.scriblio.net/>

AAI@edu¹⁵ user name and password. Tagging records in the library catalogue is done under the supervision of experts because every added tag must be approved by the librarian.¹⁶

Unfortunately, the Library itself does not use controlled vocabularies for the subject indexing. Only uncontrolled index terms are used for subject analysis of the library's collections by the librarians/subject experts.

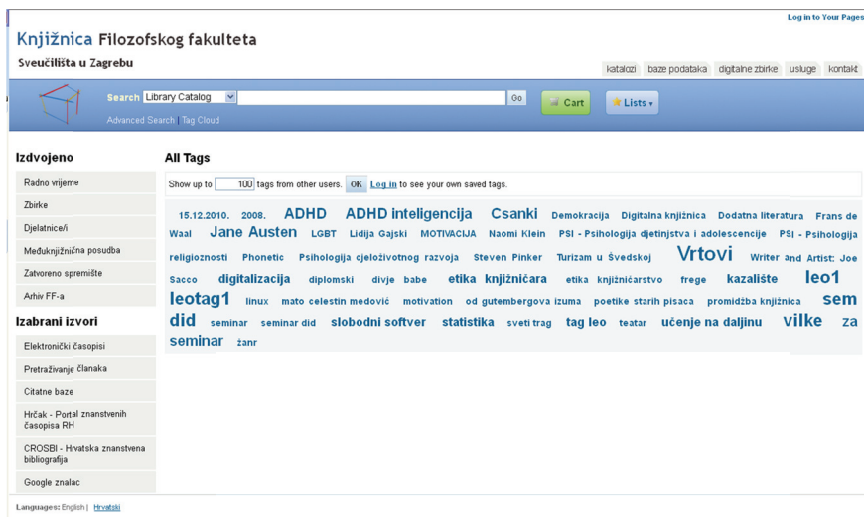


Image 1. Overview of the tag clouds in the catalogue of the Faculty of Humanities and Social Sciences Library

Research on users' behaviour in tagging systems

Numerous studies were aimed at investigating the laws regulating the distribution of tags, the type of metadata represented by the tags as well as semantic links of folksonomies to the knowledge organization system with special emphasis on analyses on users, authors, experts and computer-generated vocabularies. Research has been often conducted on the Delicious service because it is one of the largest folksonomies if we take into account the number of users, but also the number of tags on the Internet¹⁷.

¹⁵ AAI @ EduHr is the authentication and authorization infrastructure of the scientific and higher education system in the Republic of Croatia.

¹⁶ For now it is not possible to search the tags in the catalogue. Besides OPAC tagging, the Faculty of Humanities and Social Sciences Library also offers its users creating reading lists and writing comments on and about a particular library item.

¹⁷ According to Munk and Moerk, in 2007 more than 100,000 people tagged web sites that they or other people found using their key words in Delicious.

Golder i Huberman (2006) analysed the distribution of the frequencies of tags on Delicious with the aim of discovering regularities in use. The results have shown that the added key words are mostly used for the private purposes of the users, which means that the results have indicated that only that user can make cognitive connections between an individual web site and the added tags.

In late 2005, Munk and Moerk (2007) gathered 76,601 different key words from 500 randomly selected taggers on Delicious. Key words were analysed quantitatively and qualitatively. Quantitative analysis was used to analyse the frequency and the share of individual words by means of statistical correspondence analysis in order to discover possible regularities, while qualitative analysis was conducted on the textual part in order to find different strategies of tagging that the users undertake. The results of the research have shown that:

1. The distribution of tags follow the power law, which means that only a few words are dominant, and the great majority appears only once or a couple of times. The most common are the basic cognitive categories or general categories that all people or people employed in the IT sector have in common.
2. There are three tagging communities or three types of taggers: the well-informed and curious citizen who tags wide categories of culture, the IT expert who tags specific technical categories connected with IT technology and the IT designer who tags terms related to design.
3. Two hundred and forty-five most commonly used tags are situated on the axis from basic social subjects to specific concepts in the IT sector.
4. There are nine broad categories of tagging strategies; from the widest categorizing of content, categorizing of media through format, process categorizing to meta-categorization.

Kakali and Papatheodorou (2010) conducted their research on the data from the catalogue of the library of the Faculty of Social and Political Science at the Pantheon University in Athens, Greece. The researchers wanted to see why and how the students and the faculty tag bibliographic records. They were interested in the number of tags in the singular, plural, in phrases or in one word, acronyms, how many tags there are per record and how much do they correspond to expert subject headings. The results have shown that the tags are dominated by theme metadata, and that the tags for the author, title, editor and geography are a distant second. Furthermore, only 12% of tags have been identical to the subject descriptors of that record. They found that a slightly greater percentage belongs to tags consisting of two or more words, and not one word, and that the majority of tags are terms in the singular. The authors have concluded that the users use tags in order to create a short bibliography on a specific subject.

There has been a series of research conducted on the topic of implementation and tag analysis on the service LibraryThing for Libraries. LibraryThing is a

popular social tagging service used to organize personal book collection¹⁸. As of June 2011 LibraryThing had over 1,3 million users who added 76 million tags for over 63 million titles of books¹⁹. Research results have shown that as in Delicious, LibraryThing includes two broadest tag categories: meaningful and personally useful.²⁰ Lu, Park and Hu (2010) compared tags from the website LibraryThing with experts' subject descriptors in Library of Congress Subject Headings (LCSH). The authors discovered that 97.8% of all the tags cannot be found in the LCSH²¹.

This paper is the first to investigate the folksonomy of a library catalogue in Croatia. The aim of this research is to analyse what constitutes the folksonomy of the users of the Faculty for Humanities and Social Sciences Library in Zagreb, i.e. what key words the users use to organize their information in the catalogue. The research problems are as follows: which users are the most common taggers, which type of metadata is marked by tags and how do theme tags correspond to the key words in the catalogue?

Research method

The subjects of this research were the tags which were added by the users to the bibliographic records in the public online catalogue Koha of the Faculty of Humanities and Social Sciences Library. Apart from tags and categories of users who tag, we collected the following data from the bibliographic records: the bibliographic number of the record, the title, the author, co-authors, editors, key words of the subject experts and the type of library materials. The data was collected from June 15th 2010 to June 1st 2011, after which it was exported to a MS Excel table. The table was supplemented with data on the type of tag metadata and with data on the existence or non-existence of information on the subject carried by the tag in the index term field of the bibliographic record. Metadata, as well as information on the previous existence of data, were obtained through content analysis of the tags and the MARC 21 field 653 of the record on which the tag was added. The data was processed and analysed by means of descriptive statistics in the statistical programme package SPSS.

¹⁸ Westcott, Jezmyne; Chappell, Alexandra; Lebel, Candace. LibraryThing for libraries at Claremont. // *Library Hi Tech*, 27, 1(2009); 78.

¹⁹ LibraryThing. Zeitgeist. [quoted: 2011-05-15] URL: <http://www.librarything.com/zeitgeist>

²⁰ Mendes, Luiz H.; Quinonez-Skinner, Jennie; Skaggs, Danielle. Subjecting the catalog to tagging. // *Library Hi Tech*, 27, 1(2009); 32-33.

²¹ Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congress Subject Headings. // *Journal of information science*, 36, 6(2010); 770.

The results and the discussion

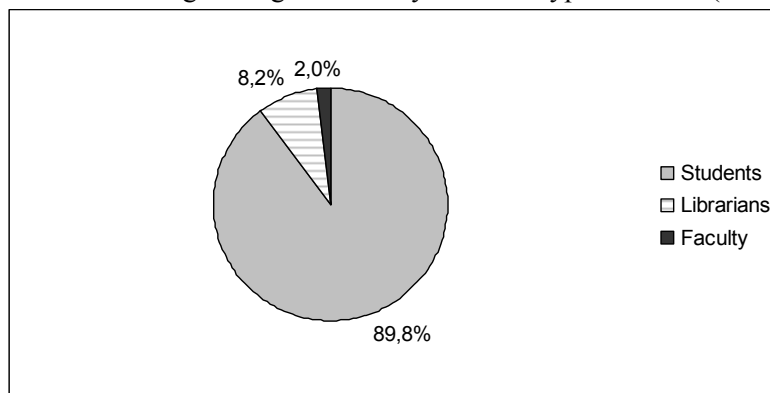
In almost a year the users added 147 tags in total to 124 bibliographic records, out of which 48 are different. The great majority of tags were added to one bibliographic record, while only 23 tags were added to 2 records (Table 1). The reason for this relatively small number of tags can be found in the probable perception of the catalogue as still belonging to the librarian-professional domain, which resulted in the users not putting in more effort to engage with the software.

Table 1. Number and percentage of tags per number of bibliographic record

	N	%
1 record	124	84.0
2 records	23	16.0
TOTAL	147	100.0

As it is shown in Chart 1, out of the three types of users, the students are the ones who tag the most. As many as 90% of tags were added by students of the Faculty of Humanities and Social Sciences in Zagreb, while librarians and the faculty have jointly added only every tenth tag. It is obvious that the students, not the academic staff, are the ones who have recognized more the usefulness of this form of knowledge organization.

Chart 1. Percentage of tags created by different types of users (N=147)



The analysis of the type of documents which were tagged by the users has shown that the books have received the highest number of tags (91.8%), and that the dissertations and theses have received only 8.2% or 12 tags. The other type of documents, such as journals or offprint has not been tagged yet. (Table 2.)

Table 2. Number and percentage of tags per type of documents

Type of documents	N	%
Books	135	91.8
Dissertations and Theses	12	8.2
TOTAL	147	100.0

Out of 147 tags, more than half of them are tags with private information (N=77). Private information means that there is no meaningful link between the name of the tag and the information in the bibliographic records. Private information is an example of what was mentioned by Golder and Huberman, and that is that such key words are of service solely to the user for private purposes and have no added value which other users could make use of. A big part of private information from the OPAC of the Faculty of Humanities and Social Sciences Library in Zagreb whose meaning can be surmised, refers to student notes on the materials that they chose for seminar or diploma papers (e.g. tags: za seminar, diplomski, sem did, seminar, seminar did, etc.). Tags carrying private information have been excluded from further analysis.

We wanted to find out which type of metadata is used by the users in order to tag bibliographic records. Content analysis revealed that the tags describe the following metadata: theme, author and the title. The most represented metadata is the theme which appears in three fourths of tags that were analysed (Table 3). Besides the theme, the metadata on the author and the title appear in smaller numbers. The results are similar to those obtained by Kakali and Papatheodouru where the theme dominated among the represented metadata.

Table 3. Number and percentage of tags per different types of metadata

Type of metadata	N	%
Theme	53	75.7
Author	8	11.4
Title	9	12.9
TOTAL	70	100.0

We further analysed only the tags with metadata on the theme since we wanted to find out how they correspond to the key words of subject experts in the bibliographic records. The results showed that in 75.5% cases, the tag of the theme carries completely new information on the subject matter which can be understood as enrichment of the bibliographic record of the tagged material (Table 4). Even though their real number is small (N=44), it can be one of the indicators of the need to revise the indexing system of the Faculty of Humanities and Social Sciences Library.

Table 4. Number and percentage of tags identical to experts' key word on the bibliographic record

Is there an identical experts' key word in the record?	N	%
No	40	75.5
Yes	13	24.5
TOTAL	53	100.0

Conclusion

This is the first research conducted in the library catalogues in Croatia on the subject of folksonomies. The research was aimed to investigate the ways in which users of The Faculty of Humanities and Social Sciences Library organize the knowledge that they need in the catalogue. The selected characteristics of the tags added to the bibliographic library records were also analysed.

The results have shown that the biggest creators of tags are the students of the Faculty of Humanities and Social Sciences and that they mainly use the tags to collect bibliography selected for writing seminar or diploma papers. Out of three types of metadata (author, title, theme), the users mainly write the theme as the tag. An interesting finding was that three fourths of theme tags are not noted in the MARC 21 field 653 of the tagged bibliographic record, meaning that they do not exist in the key words of the subject experts. Apart from this result, the need for revising the indexing system of the library is also indicated by the fact that the Faculty of Humanities and Social Sciences Library does not have a controlled vocabulary for the subject indexing of the bibliographic records.

Of course, we have to keep in mind that the number of added tags is very small and that only a significantly greater number would make it safer to make conclusions about the main tendencies in the folksonomy of the users of this library.

The tags should be taken as the user suggestions of the new terminology or an alternative to already existing concepts.

References

- Golder, Scott A.; Huberman, Bernardo. Usage patterns of collaborative tagging systems. // *Journal of Information Science*. 32, 2(2006); 198–208.
- Gordon-Murnane, Laura. Social bookmarking, folksonomies and web 2.0 tools. // *Searcher*. 14, 6(2006): 26-38.
- Kakali, Constantia; Papatheodorou, Christos. Exploitation of folksonomies in subject analysis. // *Library and information science research*. 32(2010); 192-202.
- Lu, Caimei; Park, Jung-ran; Hu, Xiaihua. User tags versus expert-assigned subject terms : a comparison of LibraryThing tags and Library of Congress Subject Headings. // *Journal of information science*, 36, 6(2010); 763-779.
- Macan, Bojan. Tehnologije Web 2.0 i njihova primjena u knjižnicama – iskustva Knjižnica Instituta „Ruder Bošković“ s posebnim osvrtom na njezin blog. // *Kemija u industriji*. 58, 5(2009); 226-228.
- Mendes, Luiz H.; Quinonez-Skinner, Jennie; Skaggs, Danielle. Subjecting tha catalog to tagging. // *Library Hi Tech*, 27, 1(2009); 30-41.

- Munk, Timme Bisgaard; Moerk, Kristian. Folksonomies, tagging communities and tagging strategies: an empirical study. // Knowledge organization. 34, 3(2007); 115-127.
- Peterson, Elaine. Beneath the Metadata : Some Philosophical Problems with Folksonomy // D-Lib Magazine. 12, 11(2006). [quoted: 2010-09-15] URL: <http://www.dlib.org/dlib/november06/peterson/11peterson.html>
- Steele, Tom. The new cooperative cataloging. // Library Hi Tech, 27, 1(2009); 68-77.
- Thomas, Marliese; Caudle, Dana M.; Schmitz, Cecilia. Trashy tags : problematic tags in LibraryThing. // New Library World, 111, 5/6(2010); 223-235.
- Westcott, Jezmyne; Chappell, Alexandra; Lebel, Candace. LibraryThing for libraries at Claremont. // Library Hi Tech, 27, 1(2009); 78-81.